

CAI3101: Introduction to Artificial Intelligence Project Description

Project Overview

In this project, students will apply machine learning techniques to analyze a real-world dataset, build predictive models, and extract meaningful insights. The goal is to demonstrate understanding of data preprocessing, model selection, evaluation metrics, and result interpretation. Students are required to submit a written report and present their findings during the discussion session.

Project Objectives

By completing this project, students should be able to:

1. Understand the end-to-end workflow of a machine learning project.
2. Perform data cleaning, preprocessing, and feature engineering.
3. Train, and compare multiple ML models.
4. Evaluate models using appropriate metrics.
5. Interpret results and translate them into actionable insights.
6. Communicate technical results in a clear academic style.

Dataset Options

Students may choose **one** of the following options:

1. **Public dataset** from Kaggle, UCI Machine Learning Repository, or Google Dataset Search.

The dataset must contain at least **200 samples** and **5 meaningful features**.

Project Tasks

1. Data Understanding

- Describe the dataset source and its context.
- Provide summary statistics and data visualizations.
- Identify the problem type (classification, regression, clustering).

2. Data Preprocessing

- Handle missing values, outliers, and inconsistent formats.
- Encode categorical variables and scale numerical ones.
- Perform feature engineering where appropriate.

3. Model Development

- Choose **at least three** machine learning models (e.g., Naïve Bais, Decision Tree, Artificial Neural Network).
- Justify the choice of algorithms.
- Split the dataset into training/testing sets.

4. Model Evaluation

- Evaluate models using suitable metrics:
 - Classification: accuracy
- Compare models and discuss results.

Submission and Discussion Criteria

- 1- Each Group must submit a report for the project and a powerpoint.
- 2- Each group consists of minimum 3 and maximum 5.
- 3- Each group must present and discuss on his time slot or there will be a **penalty** in the grades.